DOCUMENT RESUME

ED 135 938 CE 008 857

AUTHOR	Patton, Michael Q.; And Others
TITLE	In Search of Impact: An Analysis of the Utilization of Federal Health Evaluation Research.
INSTITUTION	Minnesota Univ., Minneapolis. Center for Social Research.; Minnesota Univ., Minneapolis. Dept. of Sociology.
SPONS AGENCY	Department of Health, Education, and Welfare, Washington, D.C.
PUB DATE	75
NOTE	46p.
AVAILABLE FROM	Michael Q. Patton, Director, Center for Social Research, University of Minnesota, Minneapolis, Minnesota 55455 ($2.50)
EDRS PRICE	MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS	Administrative Principles; Decision Making; Government Employees; Health Personnel; *Information Needs; Information Utilization; Problem Solving; Program Development; *Program Evaluation; Researchers; *Research Problems; *Research Utilization; Use Studies

ABSTRACT

Research on the utilizations of evaluations was based on a followup of 20 Federal health program evaluations to assess the degree to which the evaluations had been used and to identify the factors that affected varying degrees of utilization. Interviews were conducted with project officers or people they identified as decisionmakers who would utilize information in the evaluation reports. Two major themes emerged from the study. First, it was found that much of the evaluation literature has considerably overestimated the kind of impact evaluation research is likely to have. Second, the importance of the personal factor in evaluation research, particularly the utilization process, has been considerably underestimated. The two themes are directly linked. The impact of evaluation research is most often experienced as a reduction in the uncertainty faced by individual decisionmakers as they attempt to deal with the complexity of programing reality. It must be assimilated and fitted into a contextual whole. Energetic and interested people in government can and do use evaluation research, not for making decisions with immediate, concrete, and visible impacts, but in a more subtle, clarifying, reinforcing, and reorienting way. Evaluators, then, might do well to spend less time lamenting their lack of visible impact on major decisions and more of their time providing relevant information to those key persons whose thoughts and actions, to a substantial extent, determine the general direction in the evolutionary process of program development. It is in consciously working with such decisionmakers to answer their questions that the utilization of evaluation research can be enhanced. (TA)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED135938

IN SEARCH OF IMPACT:

An Analysis of the Utilization of Federal Health Evaluation Research

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Michael Patton

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Michael Q. Patton, Patricia Smith Grimes,
Kathryn M. Guthrie, Nancy J. Brennan,
Barbara Dickey French, Dale A. Blyth

CE 008 857

Copyright, 1975
Minnesota Center for Social Research
University of Minnesota
Minneapolis, Minnesota

It is 7:30 A.M. on a cold November morning. Some fifteen people in various states of wakefulness have come together to discuss a county evaluation program. The evaluation staff is there; the evaluation advisory board representatives; the county board representatives; the state representatives; and we represent the academic community.

The evaluator explains what her staff has done during the year, the problems with getting started (fuzzy program goals, uncertain funding), the data collection problems (lack of staff, little program cooperation, inconsistent state and county data processing systems), the management problems (unclear decision-making hierarchies, political undercurrents, trying to do too much), and the findings despite it all ("tentative to be sure but more than we knew a year ago and some solid recommendations").

Then the advisory board explains its frustration with the disappointing results of the evaluation program ("the data just aren't solid enough") and the county board representatives explain why their decisions are contrary to evaluation recommendations ("we didn't really get the information we needed when we wanted it and it wasn't what we wanted when we got it"). The room is filled with disappointment, frustration, defensiveness, cynicism, and more than a little anger. There are charges, counter-charges, budget threats, moments of planning and longer moments of explaining away problems. Then the advisory board chairperson turns to us--the neutral, academic observers, the evaluation experts who teach evaluation methodology at the University--and asks: "Tell us, what do we have to do to produce good evaluation research that actually gets used?"

How many times has that question been asked in the last several years? There is no shortage of advice. At program evaluation planning meetings: "Make sure your program goals are clear and operationally defined." At county, city,

state board sessions and budget committees: "Specify clearly your information needs, make clear what information you really need to make decisions." At professional academic meetings: "We need more rigorous methodologies, more sophisticated statistical analyses in evaluation research--and more funding to conduct methodologically sound evaluation research." At evaluation training sessions: "Make it clear to decision-makers what you can and can't do, be objective, maintain your integrity, work closely with relevant decision-makers, get your work done on time, and do the best job you can."

The increasing demand for evaluation research has been matched by a proliferation of books on how to do evaluation research (e.g., Weiss, 1972a and b; Suchman, 1967; Caro, 1971; American Institutes for Research, 1970; Rossi and Williams, 1972; Cherney, 1971, and Tripodi, Fellin, and Epstein, 1971). Yet decision-makers continue to lament the disappointing results of evaluation research complaining that the findings don't tell them what they need to know. And evaluators continue to complain about many things, "but their most common complaint is that their findings are ignored" (Weiss, 1972a:319).

THE ISSUE OF NON-UTILIZATION

The problem of the non-utilization or the under-utilization of evaluation research has been discussed frequently in the evaluation literature. There seems to be a consensus that the impact of evaluative research on program decision-making has been less than substantial. Carol Weiss lists under-utilization as one of the foremost problems in evaluation research:

> Evaluation research is meant for immediate and direct use in improving the quality of social programming. Yet a review of evaluation experience suggests that evaluation results have not exerted significant influence on program decisions. (Weiss, 1972a:10,11.)

Other prominent reviewers have reached a similar conclusion. Ernest House (1972:412) put it this way: "Producing data is one thing! Getting it used is quite another." Williams and Evans (1969:453) write that "in the final analysis,

the test of the effectiveness of outcome data is its impact on implemented policy. By this standard, there is a dearth of successful evaluation studies." Wholey (1971:46) concluded that "the recent literature is unanimous in announcing the general failure of evaluation to affect decision-making in a significant way." He goes on to note that his own study "found the same absence of successful evaluations noted by other authors (1971:48)." David Cohen (1975:19) finds that "there is little evidence to indicate that government planning offices have succeeded in linking social research and decision-making." Alkin (1974) found that Title VII evaluations were useful to project directors but were not useful at the Federal level because the results were not timely in terms of funding decisions. Weidman et al (1973:15) concluded that on those rare occasions when evaluation studies have been used, ". . . the little use that has occurred [has been] fortuitous rather than planned."

The variety of advice for increasing utilization that we found in the literature illustrates the complexity of the problem. It seems clear that no single prescription for success will suffice. The issue at this time is not the search for a single formula of utilization success, nor the generation of ever longer lists of possible factors affecting utilization. The task for the present is to identify and refine a few key variables that may make a major difference in a significant number of evaluation cases (cf. Weiss, 1972a:325). The research on utilization of evaluations described in this paper is a modest effort to move a bit further along that path of refinement.

This paper is based on a follow-up of twenty Federal health evaluations.[1]

---

[1]This research was conducted as part of an N.I.M.H.-supported training program in evaluation methodology at the University of Minnesota. Trainees worked through the Minnesota Center for Social Research, University of Minnesota. The following trainees, in addition to the authors, participated in the project: James Cleary, Joan Dreyer, James Fitzsimmons, Steve Froman, Kathy Gilder, David Jones, Leah Harvey, Gary Miller, Gail Nordheim, Julia Nutter, Darla Sandhoffer, Jerome Segal, and John Townsend. In addition, the following Minnesota faculty made helpful comments on an earlier draft of this paper: John Brandl, Director, School of Public Affairs;

We attempted to assess the degree to which these evaluations had been used and to identify the factors that affected varying degrees of utilization. Given the pessimistic nature of most writings on utilization we began our study fully expecting our major problem would be to find even one evaluation that had had a significant impact on program decisions. What we found was considerably more complex and less dismal than our original impressions led us to expect. Evaluation research is used but not in the ways we had anticipated. Moreover, we found that the factors we had expected would be important in explaining variations in utilization were less important than a new factor that emerged from our analysis. After reviewing our sample and methodology we shall report these findings and discuss their implications.

The Sample

The twenty case studies which constitute the sample in this paper are national health program evaluations. They were selected from among 170 evaluations on file in the Office of Health Evaluation, HEW.[2] In sampling these 170 evaluations we first eliminated studies which did not examine an existing program of some kind and studies which did not include some kind of systematic data collection so as to exclude policy pieces, think pieces, and armchair reflections from our analysis.

(footnote 1 con't.)
Martha Burt, Tom Dewar, and Ron Geizer. Neala Yount transcribed over one hundred hours of interviews with unusual diligence and care.

[2]The Office of Health Evaluation coordinates most evaluation research in the health division of HEW. In 1971 this Office designed a new record-keeping system that collected abstracts of all evaluations coming through that office. 170 evaluations were collected during the period 1971-73. This became the universe of evaluations from which we chose our final sample. As part of this project a description and analysis of the content of all 170 abstracts were made. The results of that analysis can be found in Patricia Grimes, Descriptive Analysis of 170 Health Evaluations, unpublished thesis, School of Public Affairs, University of Minnesota, 1976. We wish to express our thanks to HEW officials for their assistance throughout this research project, particularly Harry Cain, Director, Office of Policy Development and Planning, Office of the Assistant Secretary for Health, and Isadore Seeman, Director, Office of Health Evaluation, Office of the Assistant Secretary for Planning and Evaluation, DHEW.

We also eliminated studies that did not examine national programs and any studies completed before 1971 or after 1973. We did this to enable a follow-up of evaluations which were recent enough to be remembered, and at the same time, evaluations which had been completed far enough in the past to allow time for potential utilization to occur. These control variables reduced the number of abstracts from 170 to 76 and gave us a more homogenous group of abstracts consisting of (1) program evaluation studies of (2) national scope where (3) some systematic data collection was done and (4) where the study was completed no earlier than 1971 and no later than 1973.

A stratified random sample of twenty studies was then drawn from among the remaining seventy-six abstracts. Twenty was the number of studies we felt we could intensively follow-up with our limited resources. The final sample was stratified by "nature of program," as indicated in Table I, to guarantee a heterogeneous but representative group of health program evaluations.

TABLE I

Sample Stratification By Nature of Program Evaluated

Nature or Focus of the Activity which was Studied	Frequency in Sample Frame	Percent in Sample Frame	Frequency in Sample
1. Training Program	17	22.4	5
2. Program to Provide Education and General Information to Public	3	3.9	1
3. Medical Treatment and Health Care Program	29	38.2	8
4. Program of Basic Medical and Health Research and Basic Research in General	4	5.3	1
5. Establishment, Implementation and Planning of Health-Related Programs	22	28.9	5
9. Other Programs	1	1.3	0
TOTAL	76	100.0	20

For reasons of confidentiality neither the actual programs evaluated nor the titles of the evaluation studies can be reported. However, we can present a general description of the sample. The twenty cases in this study consist of four evaluations of various Community Mental Health Centers program activities, four health training programs, two national assessments of laboratory proficiency, two evaluations of Neighborhood Health Center programs, studies of two health services delivery systems programs, a training program on alcoholism, a health regulatory program, a Federal loan forgiveness program, a training workshop evaluation, and two evaluations of specialized health facilities.

Seventeen of the sample studies presented data in a way that allowed comparisons to be made between projects or activities. Sixteen included some type of quantitative analysis and eight included some type of qualitative analysis. Nine of the studies included longitudinal data. Nineteen of the studies included outcome variables and fifteen included examination of implementation variables. Nine studies were based on data from non-random samples, three sampled randomly, and eight gathered data on an entire relevant population of projects or a unique project. Eighteen of the studies concluded with recommendations or explicit judgments while two only presented data findings without making judgments.

The types of evaluation studies in the final group of twenty cases range from a three week program review carried out by a single internal evaluator to a four year evaluation that cost a million and a half dollars. Six of the cases were internal evaluations and fourteen were external.

Since it is impossible to specify the universe of evaluation research studies, it is not possible to specify the degree to which this sample of twenty cases is representative of evaluation research in general. The sample is diverse in its inclusion of a broad range of evaluations. We feel that this diversity and heterogeneity increases the meaningfulness of those patterns of utilization which actually emerged in our follow-up interviews because those patterns were

not systematically related to specific types of evaluations. The importance of this will become more evident when, after describing the nature of our data in the next section, we report our findings on utilization.

Data on Utilization: The Interviews

The first purpose of this study is to examine the nature and degree of utilization of Federal evaluation research. Ideally such a study would require exhaustive follow-up with any and all persons who did or could have used the study. With very limited resources it was possible to interview only three key informants about the utilization of each of the twenty cases in the final sample. These key informants were (1) the project officer[3] for the study, (2) the person identified by the project officer as being either the decision-maker for the program evaluated or the person most knowledgeable about the study's impact,[4] and (3) the evaluator who had major responsibility for the study.

The project officer interviews were conducted primarily to identify informants, decision-makers, and evaluators who would be interviewed about the impact and utilization of the evaluations in our sample. (The results of those interviews will not be reported in this paper.) This snowball sampling technique resulted in considerable variation in who we interviewed as the "decision-makers" in each case. Most of these government informants had been or now are Office Directors (and Deputy Directors), Division Heads, or Bureau Chiefs. Overall, these

---

[3]The term "project officer" refers to the person in the Federal government who was identified as having primary responsibility for administering the evaluation. For studies which were done by organizations which are not a part of the Federal government, the project officer was the person who administered the Federal government's contract with that organization.

[4]We identified decision-makers by asking the project officers to name a person who would serve as an informant "about how the study was used in the government or elsewhere," a person who:

> "might be called a 'decision-maker' vis-a-viz the study and its findings; who could tell us what decisions, if any, were made on the basis of information contained in the study."

decision-makers each represent an average of over fourteen years experience in the Federal government.[5]

The evaluators in our sample make for a rather heterogeneous group. Six of the twenty cases were internal evaluations so that the evaluators were Federal administrators or researchers. In one case the evaluation was contracted from one unit of the Federal government to another so that these evaluators were also Federal researchers. The remaining thirteen evaluations were conducted by private organizations or non-government employees though several persons in this group had either formerly worked in the Federal government or had since come to work directly in the government. Evaluators in our sample each represent an average of nearly fourteen years experience in conducting evaluative research.[6]

Two forms of the interview were developed--one for government decision-makers and one for evaluators. Each form was pilot-tested and revised before the final format was established. Both interviews are open-ended with questions covering the following areas: (1) Interviewee background, involvement in the program, and involvement in the evaluation; (2) purpose and objectives of the evaluation; (3) political context; (4) expectations during the study about how the findings would be used; (5) major findings from the evaluation; (6) ways in which the study had an impact on program operations, program planning, program policy, etc.; (7) non-program impacts, i.e., broader impacts on issues associated with the evaluation, position papers, new legislation, etc.; (8) impact and reception of specific study recommendations; (9) factors explaining the study's impact including specific questions on eleven factors (to be described later) taken

---

[5]In two of our twenty cases we have no information on decision-maker experience; this average is based on eighteen respondents.

[6]In four of our twenty cases we have no information on evaluator's experience; this average is based on sixteen respondents.

from the utilization literature on evaluation research; and (10) general questions on interviewee's observations about evaluation research and its utilization. The interviews were taped and transcribed, and ranged in length from one to six hours with an average of about two hours.

The analysis of the interviews began with general discussions in which the seventeen interviewers shared their perceptions about their own interviews. Three staff members then independently read all interviews looking for patterns and themes. These processes led to the formation of tentative hypotheses about dominant themes. The interview transcripts were then examined again, searching for evidence supporting these tentative hypotheses as well as looking for contradictory evidence and counter-examples. Quotes extracted from the interviews as examples of particular points were then independently examined by other staff members to check for context and accuracy. Only those findings about which there was a high degree of consensus are reported here.[7]

IMPACT OF EVALUATION RESEARCH

The conceptualization and operationalization of the notion of research impact or evaluation utilization is no easy task. We began with an ideal-typical construct of utilization as immediate and concrete effect on specific decisions and program activities resulting directly from evaluative research findings. Yet, as noted earlier, the consensus in the evaluation literature is that instances of such impact are relatively rare.

Given the dismal conclusions of most studies of utilization we began our study anticipating that our major problem would be finding even one evaluation that had had a significant and identifiable impact on program decisions. Because

---

[7]This paper represents the initial and general results of our analysis. A more extensive and detailed description of the sample, methodology, and analysis is presented in Nancy J. Brennan, Variation in the Utilization of Evaluation Research in Decision Making, unpublished Ph.D. dissertation, University of Minnesota, forthcoming.

we expected little evidence of impact and because of our inability to agree on an operational definition of utilization, we adopted an open-ended strategy in our interviewing which allowed respondents to define utilization in terms meaningful to them. Our question was as follows:

> Now we'd like to focus on the actual impact of this evaluation study. We'd like to get at any ways in which the study may have had an impact -- an impact on program operations, on planning, on funding, on policy, on decisions, on thinking about the program, and so forth.
>
> From your point of view, what was the impact of this evaluation study on the program we've been discussing?

Following a set of probes and additional questions, depending upon the respondents' initial answers, we asked a question about the non-program impacts of the evaluation:

> We've been focusing mainly on the study's impact on the program itself. Sometimes studies have a broader impact on things beyond an immediate program, things like general thinking on issues that arise from a study, or position papers, or legislation. . .
>
> Did this evaluation have an impact on any of these kinds of things?

What we found in response to these questions on impact was considerably more complex and less dismal than our original thinking had led us to expect. We found that evaluation research is used by decision-makers but not in the clear-cut and organization-shaking ways that social scientists sometimes believe research should be used. The problem we have come to feel may well lie more in many social scientists' overly grand expectations about their own importance to policy decisions than in the intransience of Federal bureaucrats. The results of our interviews suggest that what is typically characterized as under-utilization or non-utilization of evaluation research can be attributed in substantial degree to a definition of utilization that is too narrow and fails to take into consideration the nature of actual decision-making processes in most programs.

The Findings on Impact

In response to the first question on impact, fourteen of eighteen responding decision-makers and thirteen of fourteen responding evaluators felt that the evaluation had had an impact on the program. (Two of the decision-makers and six of the evaluators felt that they had too little direct knowledge of actual use to comment.) Moreover, thirteen of sixteen responding decision-makers and nine of thirteen responding evaluators felt these specific evaluation studies had had identifiable non-program impacts.

The number of positive responses to the questions on impact are quite striking considering the predominance of the theme of non-utilization in the evaluation literature. The main difference here, however, may be that the actual participants in each specific evaluation process were asked to define impact in terms that were meaningful to them and their situations. Thus, none of the impacts described was of the type where new findings from an evaluation led directly and immediately to the making of major, concrete program decisions. The more typical impact was one where the evaluation findings provided additional pieces of information in the difficult puzzle of program action permitting some reduction in the uncertainty within which any Federal decision-maker inevitably operates.

The most dramatic example of utilization reported in our sample was the case of an evaluation of a pilot program. The program administrator had been favorable to the program in principle, was uncertain what the results would be, but was "hoping the results would be positive." The evaluation proved to be negative. The administrator was "surprised, but not alarmingly so. . . We had expected a more positive finding or we would not have engaged in the pilot studies." (DM367:13)[8] The program was subsequently ended with the evaluation carrying "about a third of the weight of the total decision." (DM367:8)

---

[8]Citations for quotes taken from the interview transcripts will use the following format. (DM 367:13) refers to the transcript of an interview with

This relatively dramatic impact stood out as a clear exception to the more typical pattern where evaluation findings constitute an additional input into an on-going, evolutionary process of program action. One decision-maker with twenty-nine years experience in the Federal government, much of that time directing research, gave the following report on the impact of the evaluation study about which he was interviewed:

> It served two purposes. One is that it resolved a lot of doubts and confusions and misunderstandings that the advisory committe had. . . And the second one was that it gave me additional knowledge to support facts that I already knew, and, as I say, broadened the scope more than I realized. In other words, the perceptions of where the organization was going and what it was accomplishing were a lot worse than I had anticipated. . ., but I was somewhat startled to find out that they were worse, yet it wasn't very hard because it was partly confirming things that I was observing. (DM232:17)

He goes on to say that following the evaluation. . .

> we changed our whole functional approach to looking at the identification of what we should be working on. But again I have a hard time because these things, none of these things occurred overnight, and in an evolutionary process it's hard to say, you know, at what point it made a significant difference or what point did it merely verify and strengthen the resolve that you already had." (DM232:17)

This decision-maker had become highly involved in applied government research, including his initiation of the study in our sample, because he believed research can help reduce uncertainty in decision-making.

> As time came on I more clearly recognized that I was not satisfied with having to make program decisions that I was making or that others were making based on "professional judgment." Not that it's bad or anything,

---

(footnote 8 con't.)
a decision-maker about evaluation study number 367. The quote is taken from p. 13 of the transcript. The study numbers and page numbers have been systematically altered to protect the confidentiality of the interviewees. The study numbers do not correspond to any codes used with in DHEW. Thus (EV201:10) and (PO201:6) refer to interviews about the same study, the former was an interview with the evaluator, the latter was an interview with the project officer.

> it's just that it's pretty shaky at times, and you know, you always sit back and say, "now if I hadn't done that and done something else, what would have been the result?" So it's nice to find that there are better ways of doing it. (DM232:25)

Still his assessment of the actual impact of the evaluation was quite constrained: "It filled in the gaps and pieces that various ones really had in their orientation to the program." (DM232:12) "It verified my suspicions." (DM232:24)

Respondents frequently had difficulty assessing the degree to which an evaluation study actually affected decisions made after completion of the evaluation. This was true, for example, in the case of a large-scale evaluation effort that had been extremely expensive and had taken place over several years time. The evaluation found some deficiencies in the program, but the overall findings were quite positive. Changes corresponding to those recommended in the study occurred when the report was published, but those changes could not be directly and simply attributed to the evaluation:

> The staff was aware that the activities in the centers were deficient from other studies that we had done, and they were beefing up these budgets and providing technical assistance to some of the projects and improving mental health activities. Now I can't link this finding and that activity. Again that confirms that finding and you say, eureka, I have found _____ _____ deficient, therefore I will [change] the program. That didn't happen. [The] deficiency was previously noted. A lot of studies like this confirmed what close-by people know and they were already taking actions before the findings. So you can't link the finding to the action, that's just confirmation. . . The direct link between the finding and the program decision is very diffuse. [Its major impact was] confirming our setting, a credibility, a tone of additional credibility to the program. (DM361:12,13)

Moreover, this decision-maker felt that additional credibility for the program became one part of an overall process of information flow that helped to some degree reduce the uncertainty faced by decision-makers responsible for the program. "People in the budget channels at OMB were, I guess, eager for and interested in any data that would help them make decisions, and this was certainly one useful bit of data." (DM361:13)

The kind of impact we found, then, was that evaluation research provides some additional information that is judged and used in the context of other

available information to help reduce the unknowns in the making of difficult decisions. The impact ranges from "it sort of confirmed our impressions. . ., confirming some other anecdotal or impression that we had" (DM209:7,1) to providing a new awareness that can carry over into other programs:

> Some of our subsequent decisions on some of our other programs were probably based on information that came out of this study. . . The most significant information from this study that we really had not realized. . .made an impact on future decisions with regard to other programs that we carry on. (DM209:7)

And why did it have this impact?

> Well I guess I'll go back to the points I've already made, that it confirmed some impressionistic feelings and anecdotal information that we had about certain kinds of things. At least it gave us some hard data on which to base some future programming decisions. It may not have been the only data, but it was confirming data, and I think that's important. . . And you know at the time this study was conceived, and even by the time it was reported to us, we really had very little data, and you know, probably when you don't have any data, every little bit helps. (DM209:15)

This reduction of uncertainty emerged as highly important to decision-makers. In some cases it simply made them more confident and determined. On the other hand, where the need for change is indicated an evaluation study can help speed up the process of change or provide an impetus for finally getting things rolling.

> Well I think that all we did was probably speed up the process. I think that they were getting there anyhow. They knew that their performance was being criticized by various parts of the government and the private sector. As I said earlier, we didn't enter this study thinking that we were going to break any new ground, and when we got finished, we knew that we hadn't. All we did was document what the people have been saying for a long time--that _____ are doing a lousy job, so what else is new? But we were able to show just how poor a job they were doing. (EV268:12)

Reducing uncertainty, speeding things up, and getting things finally started are real impacts--not revolutionary, organization-shaking impacts--but real, important impacts in the opinion of the people we interviewed. One administrator summarized this view both on the specific evaluation in question and about evaluation in general as follows:

> Well, I've worn several hats. I've been on evaluation teams. I've participated in extensive evaluation in-house of other organizational components. Myself I have a favorable view toward evaluating. If nothing else it precipitates activity many times that could not be precipitated without someone taking a hard look at an organization. It did precipitate activity in [this program]. Some of it was not positive. Some of it was negative. At least something occurred that wouldn't have occurred if the evaluation hadn't taken place. (DM312:21)

Another evaluator made it quite clear that simply reducing the enormous uncertainty facing many program administrators is a major purpose of evaluative research.

> One of the things I think often is that the government itself gets scared. . .of whatever kinds of new venture that they want to go into, and they're quite uncertain as to what steps they want to take next. So then they say, okay, let's have some outside person do this for us, or maybe an inside person do this, so at least we have some "data" to base some of our policies on. (EV283:34)

The view of evaluation research that emerges in our interviews stands in stark contrast to the image of utilization that is presented as the ideal in the bulk of the evaluation literature, or at least the impression with which that literature left us. The ideal held forth in the literature we reviewed earlier is one of major impact on concrete decisions. The image that emerges in our interviews is that there are few major, direction-changing decisions in most programming, and that evaluation research is used as one piece of information that feeds into a slow, evolutionary process of program development. Program development is a process of "muddling through" (Lindblom, 1959; Allison, 1971; Steinbruner, 1974) and evaluation research is part of the muddling.

Neither did we find much expectation that government decision-making could be or should be otherwise. One person with thirty-five years experience in the Federal government (twenty of those years in evaluation) put it like this: "I don't think an evaluation's ever totally used. That was true whether I was using them as an administrator or doing them myself." (EV346:11) Later in the interview he said:

> I don't think the government should go out and use every evaluation it gets. I think sometimes just the insights of the evaluation feed over to the

> next administrative reiteration, maybe just the right way to do it. That is, [decisions aren't] clearly the result of evaluation. There's a feedback in some way. . ., upgrading or a shifting of direction because of it. [Change] it is, you know, small and slow. . ." (EV346:16)

An evaluator expressed a similar view.

> I think it's just like everything else in life, if you're at the right place at the right time, it can be useful, but it's obviously only probably one ingredient in the information process. It's rather naive and presumptuous on the part of the evaluation community and also it presumes a rationality that in no way fits. (EV264:18)

Our findings, then, suggest that the predominant image of non-utilization that characterizes much of the commentary on evaluation research can be attributed in substantial degree to a definition of utilization that is too narrow in its emphasis on seeing immediate, direct, and concrete impact on program decisions. Such a narrow definition fails to take into account the nature of most actual program development processes.

Perhaps many social scientists have come into applied government research with high hopes of rationalizing the system. Like the Peace Corps volunteers of the sixties who set off to change the world and ended up touching a few villages and a few individuals, many evaluators seem to have entered the arena of applied research expecting to make great policy waves and are disillusioned to find that they've only provided a few cogs in the great gears of program change and development, helped with a decision here or there, made actions more certain for a few decision-makers.

Yet the situation seems little different in basic research. Researchers in any field of specialization can count the studies of major impact on one hand. Most science falls into that great amorphous activity called "normal science." Changes come slowly. Individual researchers conribute a bit here and a bit there, reducing uncertainty gradually over time. Scientific revolutions are infrequent and slow in coming (Kuhn, 1972).

The situation is the same in applied research. Evaluation research is one part of the normal "science" of government decision-making. Research impacts in ripples not in waves. Occasionally a major study emerges with great impact. But most applied research can be expected to make no more than a small and momentary splash in the great pond of government. The epitaph for most studies will read something like this:

> [We expected that it would be used] but in a way of providing background information around the consequences of certain kinds of Federal decision-making options. But not necessarily in and of itself determining those decisions. In other words you might have some idea of what the consequences of the decision are, but there might be a lot of other factors you'd take into account in how you would decide. . . (DM264:8)

> You know, impact in some of these things implies that this one thing is going to affect things, [but] it's part of a total atmosphere, and in the balance of things it's contributing another bit of information about the importance of this particular process, but by no means is it the only thing entering into what's going on in a policy review like that at that time.

> [It had a particular impact in that] it contributed to the general information context of what was going on at the time, rather than in itself. . . It contributes to that background of understanding one of the policy issues, rather than resulting in one option versus another of policy being adopted. (DM264:11)

FACTORS AFFECTING UTILIZATION

We began this paper with the observation that there is no shortage of advice about how to increase the utilization of evaluation research. At the same time, it is clear that no single prescription for success will suffice. A fairly substantial list of variables or factors that may affect the degree of research impact has been generated by the literature on utilization of social science in decision-making. These explanatory factors fall into three basic categories:

1. Characteristics of the organization
   a. The constraints of decision-making in national-bureaucratic organizations (e.g., Thompson, 1967)
   b. New and innovative agencies versus older, established agencies (Weiss, 1972a)

c. Communication patterns in organizations (Mitchell, 1973; Jain, 1970.)
d. Level in the organization where evaluation is done (Alkin, 1974).
e. The degree of politicization of the organization (Weiss, 1972b).

2. Characteristics of Actors in the System: Evaluators and Decision-Makers

a. Two-Communities Theory, different values, languages, reward systems, and affiliations (Caplan et al, 1975; Halpert, 1969; Engstrom, 1970).
b. Academic versus private contract evaluators (Bernstein and Freeman, 1975).
c. Experience in doing and using research (Glock, 1961).
d. Leadership (Engstrom, 1970).
e. Internal versus external evaluators (Weiss, 1972a).
f. Reputation and legitimacy of the evaluator (Archibald, 1970).

3. Characteristics of the Evaluation

a. Methodological Quality (Bernstein and Freeman, 1975; Weiss, 1972a; Cohen, 1975; Glazer and Taylor, 1969).
b. Relationship to original objectives (Glazer and Taylor, 1969).
c. Bias (Mitchell, 1973).
d. Complexity (Glock, 1961).
e. Degree of dissemination (Halpert, 1969).
f. Relevance to agency.
g. Format of the report (Alkin, 1974: Glazer and Taylor, 1969).
h. Positive versus negative findings (Weiss, 1972a).
i. Timeliness (Mitchell, 1973).
j. Formative versus summative (Scriven, 1967; Alkin, 1974).
k. Scope of recommendations for change (Weiss, 1972a).

This list of possible factors is only meant to be suggestive not exhaustive. It is unlikely that any single study on utilization will be able to examine the importance of all such factors. The task for the present is to identify and refine a few key variables that may make a major difference in a significant number of evaluation cases (cf. Weiss, 1972a:325).

The Data on Factors Affecting Utilization

In our own research we took a dual approach to this problem of variable specification. Once the respondents had discussed their perceptions about the nature and degree of utilization of the specific evaluation study under investigation we asked the following open-ended question:

> Okay, you've described the impact of the study. Now we'd like you to think about why this study was used in the ways you've just described. Some of this you've already done, but now we'd like to explore this in more detail.

What do you feel were the important reasons why this study had the level of impact it did?

(CLARIFICATION, IF NECESSARY:)

The literature on evaluation studies suggests a lot of reasons why some evaluations are used while others are ignored. Most of the literature, however, is based on speculation. A major objective of the interviews we're doing is to find out directly from people who are in a position to know, what factors they consider important in explaining how specific studies are used.

Following a set of probes and follow-up questions, depending on the initial response to the above question, we asked respondents to comment on the relevance and importance of eleven factors extracted from the literature on utilization: methodological quality, methodological appropriateness, timeliness, lateness of report, positive/negative findings, surprise of findings, central/ peripheral program objectives evaluated, presence/absence of related studies, political factors, government-evaluator interactions, and resources available for the study. Finally, we asked respondents to "pick out the single factor you feel had the greatest effect on how this study was used."

Two related factors emerged as important in our interviews: (1) a political considerations factor and (2) a factor we have called the personal factor. This latter factor was unexpected and its clear importance to our respondents has, we believe, substantial implications for the utilization of evaluation research. <u>None of the other specific literature factors about which we asked questions emerged as important with any consistency</u>. Moreover, when these specific factors were important in explaining the utilization or non-utilization of a particular study it was virtually always in the context of a larger set of circumstances and conditions related to the issues and decisions at hand.

In the pages which follow we shall briefly review our findings with regard to the importance of these specific factors and then examine in some detail the major factor which did emerge as consistently important to an understanding of variations in utilization of evaluation research.

Lateness of Study Completion

There is a general impression from the literature cited earlier that much evaluation research is under-utilized because studies are completed too late to be used for making a specific decision, particularly budgetary decisions. This problem is based to a large extent on the notion that the purpose of evaluation research is to serve as the basis for the making of specific, identifiable, and concrete decisions. Inasmuch as we have already argued that most evaluation research does not serve such a narrow function and is not intended to serve such a narrow function, it is not surprising that lateness in the completion of studies was not an important factor in explaining utilization of the studies in our sample.

In four of our twenty cases decision-makers indicated that the final research reports were completed late, but in all four cases preliminary information was available to a sufficient extent to be used at the time the study should have been completed. In no case was lateness considered the critical factor in explaining the limited utilization of the studies. Rather, the information was viewed as feeding into a longer term process of program development and decision-making. Several decision-makers commented that it was helpful to have the information on time, but had the final report been late the impact of the study would not likely have been different. This is partly because few issues become one-time decisions. As one decision-maker put it:

> [The] study was too late for the immediate budget that it was supposed to impact on, but it wasn't too late in terms of the fact that the same issue was occuring every year after that anyway. (DM264:16)

Another decision-maker made it quite clear that the evaluation was aimed at a broader impact than the meeting of a specific deadline.

> No, this had no timetable on it as far as I know. That we had to get this done by such and such a time in order to make such a decision. I think it was part of a grand, overall plan, rather than as something specific in order to do something. I think it was in a bigger thing than for any one decision. (DM366:13)

Methodological Quality and Appropriateness

The major factor most often identified as the reason for non-utilization is the poor quality of much evaluation research. Of the fifteen decision-makers who rated methodological quality of the study about which they were interviewed, five rated the methodological quality as "high," eight said it was "medium," and only two gave the study a "low" rating. Of seventeen responding evaluators there were seven "high" ratings, six "medium" responses, and four "low" ratings. No decision-maker and only one evaluator felt that the methodology used was inappropriate for researching the question at issue.

More to the point, only four decision-makers felt that methodological quality was "very important" in explaining the study's utilization. Further probing, however, revealed that "methodological quality" meant different things to different decision-makers. For some it meant the reputation of the evaluators; for others it meant asking the right question. In no case was methodological quality identified as the most important factor explaining either utilization or non-utilization.

The relevance of methodological quality must be understood in the full context of a study, the political environment, the degree of uncertainty with which the decision-maker is faced and thus his/her relative need for any and all clarifying information. If information is scarce, then new information of even dubious quality may be somewhat helpful. For example, one administrator admitted that the evaluation's methodological rigor could be seriously questioned, but the study was highly useful in policy discussions.

> The quality and the methodology were not even considered. All that was considered was that management didn't know what was going on, the terms, the procedures, the program was foreign to their background. And they did not have expertise in it, so they were relying on somebody else who had the expertise to translate to them what was going on in terms that they would understand and what the problems were. (DM312:17)

Social scientists may lament this situation and may well feel that the methodology of evaluation research ought to be of high quality for value reasons,

i.e., because poor quality studies ought not be used. But there is little in our data to suggest that improving methodological quality in and of itself will have much effect on increasing the utilization of evaluation research. No matter how rigorous the methodology and no matter how sophisticated the statistical manipulations, evaluation research in most instances will still be only one piece of information in a complex and evolutionary process of program decision-making and development.

Again, the importance of methodological quality as a factor explaining utilization is tempered by the nature of the utilization we found. Were evaluations being used as the major piece of information in making critical one-time decisions, methodological rigor might be paramount. But where evaluation research is one part, often a small part, in a larger whole, decision-makers displayed less than burning interest in methodological quality. Indeed, methodology was most likely to be called into question if the evaluation was expected to play a central role in the making of a decision or if the results were particularly negative or surprising. One highly experienced administrator was quite explicit about this from a program point of view.

> Well, let me put it in another context. If it were negative findings programmatically we would have hit very hard on the methodology and tried to discredit it. You know, from the program standpoint. But since it was kind of positive findings, we said, 'Okay, here it is.' If anybody asked us about the methodological deficiencies we were never reluctant to tell them what we thought they were. Not many people asked. (DM361:13)

Political Factors

This last quote on methodological quality makes it clear that methodology, like everything else in evaluation research, can become partly a political question. The political nature of evaluation research has been well-documented. The decision-makers and evaluators in our sample demonstrated an acute awareness of the fact that social science research rarely produces clear-cut findings. Findings must be interpreted and interpretation is partly a political process,

a value-laden process where truth is partially a matter of whose ox is being gored.

Of the eleven specific factors about which we asked respondents to comment, political considerations were most often discussed as an important factor in explaining how study findings were used. Nine decision-makers and ten evaluators said that political considerations had affected how the study was used. In combination, at least one person interviewed in fifteen of the twenty cases felt that politics had entered into the utilization process. Nine decision-makers and seven evaluators felt that political considerations had been "very important" as a factor explaining utilization. On the other hand, nine decision-makers and five evaluators reported that political considerations played no part in the utilization process.

There is not space here to fully explore the nature and impact of these political factors. They include intra-agency and inter-agency rivalries; budgetary fights with OMB, the Administration, and Congress; power struggles between Washington administrators and local program personnel; internal debates about the purpose and/or accomplishments of pet programs. Budgetary battles seemed to be the most political. One evaluator was particularly adamant about the political nature of his evaluation from the initiation of the study to the final report: "This was a really hot political issue, and I think the political aspects were developing and changing, and I think that was the really important factor [explaining utilization]." (EV264:17) The decision-maker concurred.

We did not find, however, that political factors suddenly and unexpectedly surfaced once a study was completed. In almost every case both the decision-makers and evaluators were well aware of the political context at the outset. Moreover, our respondents seemed to feel political awareness on the part of everyone involved was the best one could expect. Social scientists will not change the political nature of the world, and while several respondents were quite cynical on this point, the more predominant view seemed to be that government would not

be government without politics. One particularly articulate decision-maker expressed this view quite explicitly:

> This is not a cynical statement. . . A substantial number of people have an improper concept of how politics works and what its mission is. And its mission is not to make logical decisions, unfortunately for those of us who think program considerations are important. Its mission is to detect the will of the governed group and express that will in some type of legislation or government action. And that will is very rarely, when it's pooled nationally, a rational will. It will have moral and ethical overtones, or have all kinds of emotional loads. . .
>
> It's not rational in the sense that a good scientific study would allow you to sit down and plan everybody's life, and I'm glad it's not, by the way. Because I would be very tired very early of something that ran only by the numbers. Somebody'd forget part of the numbers, so I'm not fighting the system, but I am saying that you have to be careful of what you expect from a rational study when you insert it into the system. It has a tremendous impact. . . It is a political, not a rational process. . . Life is not a very simple thing. (DM328:18-19)

The importance of political considerations in much (though clearly not all) evaluation research can be partly understood in terms of our emphasis on the role of evaluation in reducing uncertainty for decision-makers. Several organizational theorists (e.g. Thompson, 1967; Crozier, 1964) have come to view power and relationships within and between organizations as a matter of gaining control through the reduction of uncertainty. French sociologist Michel Crozier has summarized this view as follows:

> In such a context, the power of A over B depends on A's ability to predict B's behavior and on the uncertainty of B about A's behavior. As long as the requirements of action create situations of uncertainty, the individuals who have to face them have power over those who are affected by the results of their choice (1964:158).

More directly, James Thompson (1967) describes evaluation research as one major organizational mechanism for reducing internal as well as environmental uncertainty. He argues that the methodological design of much evaluation research can be predicted directly from the political function that assessment plays. We believe that our data directly supports this viewpoint. Evaluations are undertaken as a mechanism for helping decision-makers cope with the complexity of the programs

for which they have responsibility. As one weapon or tool in the struggle to gain control over organizational and program processes evaluation research can fully be expected to take on a political character. Indeed, as Thompson argues, it is completely rational for decision-makers to use evaluations in a political fashion for control and reduction of uncertainty.

It would appear to us that it behooves social scientists to inform themselves fully about the political context of the evaluations on which they work. It is precisely through such a heightened awareness of the political implications and consequences of their research that social scientists can reduce their own uncertainty about the uses to which their work is put without impairing their ability to state their "truth" as they see it.

Other Factors Affecting Utilization

None of the other factors about which we asked specific questions emerged as consistently important in explaining utilization. When these other factors were important their importance stemmed directly from the particular circumstances surrounding that evaluation and its purpose, particularly its political purpose. For example, the amount of resources devoted to a study might add to the credibility and clout of a study but more costly evaluations did not show any discernible patterns of utilization different from less costly evaluation. The resources available for the study were judged inadequate for the task at hand by only two decision-makers and five evaluators.

Whether or not findings were positive or negative had no demonstrable effect on utilization. We had studies in our sample in which the findings were rated by respondents as predominantly negative; other studies were predominantly positive in their conclusions; and still others had mixed findings. This variation was evenly distributed in our sample. Interestingly enough, the decision-maker and evaluator on the same study often differed on whether findings were "positive" or "negative," but despite such disagreements neither rated the positive or negative nature of the findings as particularly important in explaining either utilization or non-utilization of the evaluation.

The unimportance of this factor in explaining variations in utilization is related to the fact that, as noted earlier, the positive or negative findings of a particular study constitute only one piece of information that feeds into a larger process in such a way that no single study is likely to have a dramatic impact on a program. Negative or positive findings are thus interpreted in the larger context of other available information.

Furthermore, the negative or positive nature of an evaluation report was unimportant as a factor explaining utilization because such findings, in either direction, were virtually never surprising. Only four decision-makers expressed surprise at the findings of the study. Only one decision-maker felt this surprise

had an important effect on utilization. There was considerable consensus that surprises are not well received. Surprises are more likely to increase uncertainty rather than reduce uncertainty.

One decision-maker took this notion a step farther and made the point that a "good" evaluation process should build-in feedback mechanisms that guarantee the relative predictability of the content of the final report.

> If you're a good evaluator you don't want surprises. The last thing in the world you want to do is surprise people, because the ... chances are surprises are not going to be well received ... It isn't a birthday party, and people aren't really looking for surprises. So that if you're coming up with data that is different than the conventional wisdom, you ought, a good evaluation effort I would suggest, would get those ideas floated during the evaluation process so that when the final report comes out, they aren't a surprise. So my reaction was that if you were dealing in the world of surprises you aren't doing a very good evaluation. Now you could come up with findings contrary to the conventional wisdom, but you ought to be sharing those ideas, if you will, with the people being evaluated during the evaluation process to be sure that those surprises don't have any relationship to reality and again working during that process on the acceptance that maybe..., but if you present a surprise, it will tend to get rejected. See, we don't want surprises. We don't like surprises around here. (DM346:30-31)

The evaluator for this project expressed the same opinion: "Good managers are rarely surprised by the findings. If there's a surprising finding it should be rare. I mean, <u>everybody's missed this insight except this great evaluator? Nonsense!</u>" (EV364:13)

Surprises were occasionally helpful if they related to relatively minor aspects of the program where fine-tuning could take care of the problem. But there was a marked emphasis throughout the interviews on the importance of evaluation as a way of "reinforcing" already existing information, adding "credibility" and confirmation to existing or anticipated program directions, and essentially making the already suspected more certain.

Surprises may have been especially minimal in our sample because most of the studies examined central rather than minor or peripheral program objectives.

It is sometimes suggested that evaluations aren't used because they concentrate on minor issues. But only two decision-makers and one evaluator felt that the evaluation in question dealt with peripheral program objectives. On the other hand nine decision-makers and eight evaluators felt that a major factor in utilization was whether or not the evaluation examined central program objectives. The most useful evaluations were those that focused on central objectives but these were also precisely the kinds of evaluations that would not produce information that in and of itself could change a major policy direction. This may have been because by focusing on major objectives the studies in our sample became one part in the larger policy process while at the same time reducing the potential for the more immediate, concrete impact that a study might have had if it had been aimed at some peripheral, easily-changed policy objective or program component. We lacked sufficient cases of the latter type, however, to explore this possibility more fully.

Another factor of interest to us concerned the point in the life of the program when the evaluation took place. The interview question read as follows:

> The next factor concerns the point in the life of a program when an evaluation takes place. For example, some evaluations are done so early in the life of a program that the program is still changing a great deal and so the evaluation is already out-of-date by the time it is finished. Other evaluations are done so late in the life of a program that the decision to terminate has already been made by the time the evaluation is done. These examples represent the extremes, but at what point in life of your program did this evaluation take place?

Our sample contained studies that were done at all stages in the lives of programs. The key point that emerged with regard to this factor was that different questions emerge at different points in the life of a program. Early in the program the most useful information concerns procedures and implementation. Outcomes only become important after the program has been operating for a resonable period. Budget and cost issues become central late in the program's life. Our respondents generally felt that in each case the questions examined had been appropriate to the

point in the life of the program when the evaluation had taken place. The point in the program's history when the evaluation occurred was not a factor in explaining utilization in our data. There was no systematic relationship between this factor and degree of evaluation utilization.

A factor that did emerge as somewhat important was the presence or absence of other studies on the same issue. Studies that broke new ground were particularly helpful because their potential for reducing uncertainty was greater. Nevertheless, such studies were viewed with some caution because our decision-makers clearly favored the accumulation of as much information from as many sources as possible. Thus, those studies which could be related to other studies had a clear cumulative impact. On the whole, however, studies that broke new ground appeared to have somewhat greater identifiable impact. About half of the studies in our sample were of this latter type.

Finally, we asked our respondents about evaluator-government interactions. The studies in our sample appeared to have been based on considerable interaction. Interactions were almost universally described as cooperative, helpful, and frequent. Many respondents could offer horror stories about poor interactions on other studies, but with regard to the specific study on which we were conducting follow-up, there appeared to be few problems. Most evaluators and project officers reported that they had interacted regularly. Moreover, they rated their interactions around these studies as at least average and often above average compared to their other evaluation experiences. There was no indication that utilization would have been increased by greater government-evaluator interaction than that which actually occurred, though the degree of interaction which did occur was considered quite important.

Overview of Findings on Factors

We have reviewed eleven factors frequently identified in the evaluation literature as affecting utilization. Lateness of completion was not a problem in our

sample partly because of the nature of the impacts we indentified. Methodological quality did not emerge as an issue in the utilization of the studies we examined and respondents generally felt that the methodology employed was appropriate to the evaluation questions asked. Resources were not a problem. The content of final reports, i.e., negative or positive findings, was not predictive of utilization. Most of our sample evaluations looked at central program objectives and revealed few surprises about how programs were operating. The questions asked in these studies were generally appropriate to the point in the life of the program when the evaluation took place; and this factor too was unrelated to utilization in any systematic way. Studies that broke new ground appeared to have somewhat greater impact in reducing uncertainty in decision-making though the cumulative effect of evaluations that related to other studies emerged as important. Interaction between government administrators and evaluators was generally positive and occurred with marked frequency.

Only the importance of political considerations emerged clearly as a major factor in the utilization process. Our study confirms the conventional wisdom that evaluation research can be highly political in nature--and in use. But our respondents felt that politics could work to either increase or decrease utilization depending upon particular circumstances. Political considerations are a normal and probably inevitable factor in the use of any information that can reduce uncertainty and thereby affect power relationships within and between organizations. By being aware of political considerations evaluators can reduce their own uncertainty about the utilization of their work.

None of these factors, at least as we were able to explore them, helped us a great deal in explaining variations in utilization. This is partly because, as noted earlier, we had set out with a different concept of utilization than that which emerged from the data. Nevertheless, there was one major factor that did consistently arise in the comments of decision-makers, evaluators, and project officers--a factor so crucial that respondents repeatedly pointed to it as the

single most important element in the utilization process.

What we are about to discuss will come as no surprise to our generous respondents and informants. Yet, we did not anticipate the importance of this factor. Perhaps we can fault the structural emphasis in much of sociology; or perhaps the problem lies in the evaluation literature, in the reification of rationality and objectivity as the links between research findings and their utilization. Whatever the source of our initial narrow vision, we believe that these findings have profound implications for evaluative research and its utilization.

What then is this factor that has emerged with such striking clarity? For lack of a better term, we have called it simply the personal factor. It is made up of equal parts of leadership, interest, enthusiasm, determination, commitment, aggressiveness, and caring. Where the personal factor emerges, evaluations have an impact; where it is absent there is a marked absence of impact.

Social scientists do not generally feel very comfortable with such personal factors. They smack too much of the great person theory of history. And so as soon as the personal factor emerges the social scientist turns away, presumably to search for the underlying structural conditions that give rise to this irritating personal phenomenon. And perhaps eventually such underlying conditions will be fully indentified and operationalized. But for the monent we want to look directly at the factor itself--the effect of individual people in a system where individuals are supposed to be interchangeable in organizational roles and positions, but where they aren't interchangeable at all.

The personal factor emerged most dramatically in our interviews when, having asked respondents to comment on the importance of each of our eleven utilization factors, we asked them to identify the single factor that was most important in explaining the impact or lack of impact of that particular study. Time after time the factor they identified was not on our list. Rather, they responded in terms

of the importance of individual people.

> I would rank as the most important factor this division director's interest, [his] interest in evaluation. Not all managers are that motivated toward evaluation. (DM353:17)

> [The single most important factor that had the greatest effect on how the study got used was] the principal investigator...If I have to pick a single factor, I'll pick people any time. (DM328:20)

> That it came from the Office of the Director--that's the most important factor...The proposal came from the Office of the Director. It had had his attention and he was interested in it, and he implemented many of the things. (DM312:21)

> [The single most important factor was that] the people at the same level of decision-making in [the new office] were not interested in making decisions of the kind that the people [in the old office] were, I think that probably had the greatest impact. The fact that there was no one at [the new office] after the transfer who was making programmatic decisions. (EV361:27)

> Well, I think the answer there is in the qualities of the people for whom it was made. That's sort of a trite answer, but it's true. That's the single most important factor in any study now that's utilized. (EV232:22)

> Probably the single factor that had the greatest effect on how it was used was th insistence of the person repsonsible for the initiating the study ...at the Director of _____become familiar with its findings and arrive at a judgment on it. (DM369:25)

> [The most important factor was] the real involvement of the top decision-makers in the conceptualization and design of the study, and their commitment to the study. (DM268:9)

While these comments concern the importance of interested and committed individuals in studies that were actually used, studies that were <u>not</u> used stand out in that there was often a clear absence of the personal factor. One evaluator, who was not sure about how his study was used but suspected it had not been used, remarked: "I think that since the client wasn't terribly interested ... and the whole issue had shifted to other topics, and since we weren't interested in doing it from a research point of view ..., nobody was interested." (EV264:14)

Another evaluator was particularly adamant and articulate on the theory that the major factor affecting utilization is the personal energy, interests, abilities, and contacts of specific individuals. This person had had thirty-five years experience

in government, twenty of those years directly involved in research and evaluation. He had also worked for several years as a private evaluation research contractor during which time he had been involved in evaluating some eighty projects for HEW. Throughout his responses to our questions on the importance of various specific factors in affecting utilization he returned to the theme of individual actions. When asked to identify the one factor that is most important in whether a study gets used he summarized his viewpoint:

> The most important factor is desire on the part of the managers, both the central federal managers and the site managers. I don't think there's [any doubt], you know, that evaluation should be responsive to their needs, and if they have a real desire to get on with whatever it is they're supposed to do, they'll apply it. And if the evaluations don't meet their needs they won't. About as simple as you can get it. I think the whole process is far more dependent on the skills of the people who use it than it is on the sort of peripheral issues of politics, resources... Institutions are tough as hell to change. You can't change an institution by coming and doing an evaluation with a halo. Institutions are changed by people, in time, with a constant plugging away at the purpose you want to accomplish. And if you don't watch out, it slides back. (EV346:15-16)

His view had emerged early in the interview when he described how evaluations were used in OEO.

> In OEO it depended on who the program officer was, on the program review officials, on program monitors for each of these grant programs ... Where they were aggressive program people, they used these evaluations whether they understood them or not. They used them to affect improvements, different allocations of funds within the program, explanations of why the records were kept this way, why the reports weren't complete or whatever. Where they, where the program officials in OEO were unaggressive, passive---nothing!
>
> Same thing's true at the project level. Where you had a program director who was aggressive and understood what the hell the structure was internally, and he used it as leverage to change what went on within his program. Those who weren't--nothing. (EV346:5)

Nor did he view this emphasis on the individual as meaning evaluation was simply a political tool. When asked how political considerations affected evaluations, he replied:

> I don't think it's political at all. Oh, there's some pressures every once in a while to try and get more efficient, more money attributes, but I don't think that's the main course. The basic thing is how the administrators of the program view themselves, their responsibilities. That's the controlling factor. I don't think it's political in any way. (EV346:8)

Later he commented:

> It always falls back to the view of the administrator and his view of where his perogatives are, his responsibilities. A good manager can manage with or without evaluations and a poor one can't, with or without evaluations. It just gives him some insights into what he should or shouldn't be doing, if he's a good manager. If they're poor managers, well ...(EV346:11)

On his comments about each possible factor the same theme emerged. Methodological quality, positive or negative findings, the degree to which the findings were expected--he always eventually returned to the themes of managerial interest, competence, confidence.

The good manager is aggressive, open, confident, anxious to interchange ideas. He's not defensive. Rather, "he's interested in finding out what your views are, not defending his....You know my sample is relatively small, but I'd say propably there are a quarter (25%) of what I'd call good managers..." (EV346:15) <u>These, he believes, are the people who use evaluation research.</u>

Our sample includes another rather adamant articulation of this premise. An evaluation of a pilot program involving four major projects was undertaken at the instigation of the program administrator. He made a special effort to make sure that his question (Were the pilot projects capable of being extended and generalized?) was answered. He guaranteed this by personally taking an active interest in all parts of the study. The administrator had been favorable to the program in principle, was uncertain what the results would be, but was "hoping the results would be positive." The evaluation proved to be negative. The administrator was "surprised, but not alarmingly so...We had expected a more positive finding or we would not have engaged in the pilot studies." (DM367:13) The program was thus ended with the evaluation carrying "about a third of the weight of the total decision." (DM367:8)

The evaluator interview on this case completely substantiated the administrator's description. The findings were specific and clear. The program was not refunded.

And thus the evaluation had a substantial, direct impact on that decision. The question then becomes why this study had such significant utilization. The answer from the decision-maker was brief and to the point:

> Well, [the evaluation had an impact] because we designed the project with an evaluation component in it, so we were expected to use it and we did .. Not just the fact that [evaluation] was built in, but the fact that we built it in on purpose. This is, the agency head and myself had broad responsibilities for this, wanted the evaluation study results and we expected to use them. Therefore they were used. That's my point. If someone else had built it in because they thought it was needed, and we didn't care, I'm sure the use of the study results would have been different. (DM367:12)

As the decision-maker tells the story it had taken a great deal of direct interaction to be sure that the right question was evaluated.

> "The initial design stages went round and round because they [the evaluators] kept trying to answer a different question than the one we wanted answered ...If we had dropped it with them right then and said go ahead and do your own thing with it, it would not have been useful...I have a feeling I'm becoming redundant. The greatest single factor [explaining utilization] was that the question we wanted answered was the question they did at least try to answer in the study." (DM367:16)

The evaluator, (an external agent selected through an open RFP process), completely agreed that:

> "The principle reason [for utilization] was because the decision-maker was the guy who requested the evaluation and who used its results. That is, the organizational distance between the policy maker and the evaluator was almost zero in this instance. That's the most important reason it had an impact." (EV367:12)

> Well, I guess the point is that the project was really monitored by the decision-maker rather than the project officer...It was the fact that the guy who was asking the question was the guy who was going to make use of the answer... So it might be interesting to experiment with the idea of having the guy who needs the the answer to the question actually run the contract. Might be interesting. Might be chaos.

An initial problem arose because the project officer had written the RFP. As the evaluator recalled the situation, the RFP was highly misleading.

> If I had done exactly what the RFP asked for and turned in a report that was responsive to the RFP but not to what I was very clear were the kinds of questions they wanted answered, they would have gotten a different report. As a matter of fact, let me just tell you the essence of the thing. I had almost no direction from the government, as I've said, except that the guy kept saying, well here on point 8, you've got to do 8 on the contract.

> So when I turned in the draft of the report, I did points 1 through 9 and put that in the final report. Then I essentially wrote another report after that and made that the last half of the report. It was a detailed description of the activities of the program, it came to very specific kinds of conclusions. It wasn't what was asked for in the RFP, but it was what they needed to answer the question in the RFP. [The decision-maker] read it and the comment back was, "It's a good report, except for all that crap in the front."
>
> Okay, so I turned it around then in the final draft, and put all that crap in the front into an appendix. And if you look at the report, it has a big, several appendices. All of that, if you compare that carefully to the contract, all that crap in the appendix is what I was asked to get. All the stuff that constitutes the body of the report was above and beyond the call. (EV367:12)

What emerges here is a picture of a decision-maker who knew what information he wanted, an evaluator committed to answering the decision-maker's question, and a decision-maker committed to using that information. The result was a high level of utilization in making a decision contrary to the decision-maker's initial personal hopes. And in the words of the evaluator, the major factor explaining utilization was

> "that the guy who's going to be making the decision is aware of and interested in the findings of the study and has some hand in designing the questions to be answered, that's a very important point." (EV367:20)

The decision-maker's conclusion is so similar that it sounds like collusion.

> Evaluation research. Well I guess I would affirm that in many cases it has no impact for many of the reasons that the literature has suggested. But if I were to pick out factors that made a positive contribution to its use, one would be that the decision-makers themselves wanted the evaluation study results. I've said that several times. If that is not present, it is not surprising that the results aren't used. (DM267:17)

This point was made often in the interviews. One highly placed and highly experienced administrator offered the following advice at the end of a four hour interview:

> Win over the program people. Make sure you're hooked into the person who's going to make the decision in six months from the time you're doing the study, and make sure that he feels it's his study, that these are his ideas, and that it's focused on his values....I'm sure it enters into personality things.... (DM283:40)

The personal factor applied not just to utilization but to the whole evaluation process. Several of the studies in our sample were initiated completely by a single person because of his personal interests and information needs. One study in particular stands out because it was initiated by a new office director with no support internally and considerable opposition from other affected agencies. The director found an interested and committed evaluator. The two worked closely together. The findings were initially ignored because there was no political heat at the time, but over the ensuing four years the director and evaluator worked personally to get the attention of key Congressmen. They were finally successful in using personal contacts. The evaluation contributed to the eventual passing of significant legislation in a new area of federal control. From beginning to end the story was one of personal human effort to get evaluation results used.

The specifics vary from study to study but the pattern is markedly clear: Where the personal human factor emerges, where some individual takes direct, personal responsibility for getting the information to the right people, evaluations have an impact. Where the personal factor is absent, there is a marked absence of impact. Utilization is not simply determined by some configuration of abstract factors; it is determined in large part by real, live, caring human beings.

Implications of the Personal Factor in Evaluation

If, indeed, utilization is to a large extent dependent upon the interests, capabilities, and initiative of individuals, then there are some profound implications for evaluators. First, evaluators who care about seeing their results utilized must take more seriously their responsibility for identifying relevant decision-makers. Relevancy in the context of the personal factor means finding decision-makers who have a genuine interest in evaluation information -- persons

who know what questions they want answered and who know how they can use evaluation information once findings are available. Such individuals are willing to take the time and effort to interact with evaluators about their information needs and interests.

Secondly, formal position and authority are only partial guides in identifying relevant decision-makers. Evaluators must find a strategically located person (or persons) who is enthusiastic, committed, competent, interested, and aggressive. Our data suggest that more may be accomplished by working with a lower level person displaying these characteristics than in working with a passive, disinterested person in a higher position.

Third, regardless of what an RFP calls for, the most valuable information with the highest potential for utilization is that information that directly answers the questions of the individual(s) identified as the relevant decision-maker(s). Requests for proposals (RFPs) may be written by individuals other than the decision-makers who really need and want the evaluation information. It behooves evaluators to clarify the degree to which an RFP fully reflects the information needs of interested government officials.

Fourth, attention to the personal factor may assist not only evaluators in their efforts to increase the utilization of their research, but attention to the personal factor can also aid decision-makers in their effort to find evaluators who will provide them with relevant and useful information. Evaluators who are interested in and knowledgeable about what they're doing, and evaluators who are committed to seeing their findings utilized in answering decision-maker's questions will provide the most useful information to decision-makers.

Fifth, there are political implications for both evaluators and decision-makers in explicitly recognizing and acting on the importance of the personal factor. To do so is also to accept the assumption that decision-making in

government is likely to continue to be a largely personal and political process rather than a rationalized and scientific process. This assumption means that neither the decision-maker nor the evaluator is merely a technician at any stage in the evaluation process. The personal factor is important from initiation of the study through design and data collection stages as well as in the final report and dissemination parts of the process. If decision-makers have shown little interest in the study in its earlier stages our data suggest that they are not likely to suddenly show an interest in using the findings at the end of the study. Utilization considerations are important throughout a study not just at the stage where study findings are disseminated.

Finally, the importance of the personal factor suggests that one of the major contributing reasons for under-utilization of evaluation research is the high degree of instability in Federal program operations. This instability, based on our data, is of three kinds: (1) high turnover rates among senior government staff so that the person initially interested in an evaluation may be in an entirely different office before the study is completed; (2) reorganization of government offices so that decision-making patterns are unstable, personnel are frequently rearranged, and responsibilities are almost constantly changing; and (3) program mobility as programs move from office to office (e.g., OEO to HEW) even if no formal, structural reorganization occurs.

We found the instability of Federal organizational charts and the mobility/turnover among staff to be substantial. In trying to retrace the history of evaluations we frequently got a response like the following: "I've had so many changes in organizational assignments since then, I don't remember." (EV201:6) Asked about utilization of evaluation the same person responded:

> Well since you're not going to identify me and my name, I'll tell you what I really think, and that is, I think these plans go up to the planning office and the rotation of personnel up there in the _____'s

> office is so fast and so furious, that they never get a chance to react to them. [It] just sits. We know that happens sometimes, because the guy who asked for it is gone by the time it gets up there. (EV201:8)

The problem of instability appears to be particularly critical in actually implementing recommended changes.

> It was easier to get recommendations through with senior management approval. I mean, they read it and they could easily implement some of the areas, and they, I mean, at least they could implement them in theory anyhow. But still the problem, in any study or anything of this caliber, it's up to the people in the operating unit to make the change. And there's no way for senior management to measure that change. There's no way to see that it was even done. You know it's the old thing, of, you know, they tell people to do things, but in areas that require technical expertise there's no way to see that the change was done. And so people in the operating area many times would just wait out the person, you know, some of these people have been through 3 directors, 5 associate directors, you know, and, they don't want to do something. They have tenure, and they know that if they sit long enough that that person will pass and someone else will come in with brand new ideas and. . . (DM312:15)

Another evaluator found the same problem in the field. "I have spoken with project directors who tell me they really have no one to talk to because at the regional level the project officers and program officers change jobs so frequently that there's no continuity." (EV346:7) Evaluators commented that it was common experience to go through several project officers on an evaluation.

Our own experience in trying to locate the respondents in the sample gave us a clear indication of this instability. Few of our interviewees were still in the same office at the time of the interview that they had been in at the time of the evaluation two to three years earlier. We still haven't been able to construct a meaningful organizational chart of HEW locating the various office changes and agency reorganizations we encountered.

These structural conditions of mobility and instability make application of the personal factor in locating relevant decision-makers or evaluators a risky business. That key person you locate may be gone by the time the study is completed. Yet these same structural conditions of mobility and instability may

well be the underlying reasons why the decision-making process in the Federal government has been and continues to be a highly personal and political process.

Conclusion

Two major themes emerge from this study of the utilization of evaluation research. First, we found that much of the evaluation literature has considerably overestimated the kind of impact evaluation research is likely to have. Second, the importance of the personal factor in evaluation research, particularly the utilization process, has been considerably underestimated.

The two themes are directly linked. The impact of evaluation research is most often experienced as a reduction in the uncertainty faced by individual decision-makers as they attempt to deal with the complexity of programming reality. Evaluation information is one piece of data available to decision-makers. It must be assimilated and fitted into a contextual whole. "The results are never self-explanatory." (EV209:9) The translation, the interpretation, the meaning, the relevance--these things are established through the interactions over time of individuals who care enough to take the time to make the contextual fit, and then are interested enough to act on the basis of that contextual fit.

It is an energy-consuming process. Energetic and interested people in government can and do use evaluation research, not for the making of grand decisions with immediate, concrete, and visible impacts, but in a more subtle, clarifying, reinforcing, and reorienting way. Evaluators, then, might do well to spend less time lamenting their lack of visible impact on major decisions and more of their time providing relevant information to those key persons of energy and vision whose thoughts and actions, to a substantial extent, determine the general direction in the evolutionary process of program development. It is in consciously working with such decision-makers to answer their questions that the utilization of evaluation research can be enhanced.

References

Alkin, Marvin C., Jacqueline Kosecoff, Carol Fitz, Gibbon and Richard Seligman. Evaluation and Decision-Making: The Title VII Experience. Los Angeles: Center for the Study of Evaluation, University of California, 1974.

Allison, Graham T. Essence of Decision: Explaining the Cuban Missile Crisis. Boston: Little-Brown, 1971.

American Institutes for Research. Evaluative Research Strategies and Methods. Pittsburgh: American Institutes for Research, 1970.

Archibald, Kathleen. "Alternative Orientations to Social Science Utilization." Social Science Information, vol. 9, no. 2 (1970), 7-34.

Baizerman, Michael. "Evaluation Research and Evaluation: Scientific Social Reform Movement and Ideology." Journal of Sociology and Social Welfare, (Winter, 1974), 277-288.

Bernstein, Ilene N. and Howard W. Freeman. Academic and Entrepreneurial Research. New York: Russell Sage Foundation, 1975.

Brennan, Nancy J. Variation in the Utilization of Evaluation Research in Decision-Making. School of Social Work, University of Minnesota, Unpublished thesis, 1976.

Caplan, Nathan, Andrea Morrison and Russell J. Stambough. "The Use of Social Science Knowledge in Policy Decisions at the National Level." Ann Arbor, Michigan: Center for Research on Utilization of Scientific Knowledge. Institute for Social Research, University of Michigan, 1975.

Caro, Francis G. (ed.) Readings on Evaluative Research. New York: Russell Sage Foundation, 1971.

Cherney, Paul R. (ed.) Making Evaluation Research Useful. Columbia, Maryland: American City Corporation, 1971.

Cohen, David K., and Michael S. Garet. "Reforming Educational Policy with Applied Social Research." Harvard Educational Review, vol. 45, no. 1 (February, 1975), 17-41.

Coleman, James C. "Policy Research in the Social Sciences." General Learning Press, 1972.

Crozier, Michael. The Bureaucratic Phenomenon. Chicago: University of Chicago Press, 1974.

Engstrom, George A. "Research Utilization: The Challenge of Applying SRS Research." Welfare in Review, vol. 8, no. 5, (Sept.-Oct., 1970), 1-7.

Glaser, Edward M. "Utilization of Applicable Research and Demonstration Results." Human Interaction Research Institute. Los Angeles: Final Report to Vocational Rehabilitation Administration, HEW, (March, 1967).

Glaser, Edward M., and Samuel H. Taylor. "Factors Influencing the Success of Applied Research: A Study of Ten NIMH Funded Projects." Human Interaction Research Institute, (Jan. 1969).

Glock, Charles Y. "Applied Social Research: Some Conditions Affecting Its Utilization" in Studies in the Utilization of Behavior Science. Institute for Communication Research, 1961.

Grimes, Patricia Smith. Descriptive Analysis of 170 Health Evaluations. School of Public Affairs, University of Minnesota, Unpublished thesis, 1976.

Halpert, Harold P. "Communications as a Basic Tool in Promoting Utilization of Research Findings" in Herbert C. Schulberg, Alan Sheldon, Frank Baker (eds.). Program Evaluation in the Health Fields. New York: Behavioral Publications, 1969.

House, Ernest R. "The Conscience of Educational Evaluation." Teachers College Record, vol. 73, no. 3, (1972), 405-414.

Jain, Nemi C. "Communication Patterns and Effectiveness of Linkers in a Formal Organization," a paper presented at the Speech Communication Association Convention. New Orleans, Lousiana, (December 27-30, 1970).

Kuhn, Thomas. Structure of Scientific Revolutions. Chicago: University of Chicago Press, 1970.

Lindblom, Charles E. "The Science of Muddling Through." Public Administration Review, (Spring, 1959), 79-89.

Mitchell, Terrence R. "Process Problems with Contract Evaluations." University of Washington, Unpublished paper, 1973.

Mitchell, Terrence R. Above paper summarized in "Headway." Evaluation: A Forum for Human Service Decision Makers, vol. 1, no. 2, (1973), 21-23.

Morris, Robert S. and Martin Rein. "The Evaluation of Broad-Aim Programs: Experimental Design, Its Difficulties and an Alternative." Administrative Science Quarterly, (March 15, 1970).

Rossi, Peter H., and Walter Williams. Evaluating Social Programs: Theory, Practice and Politics. New York: Seminar Press, 1972.

Scriven, Hichael. "The Methodology of Evaluation." In Ralph W. Tyler, Robert M. Gagne and Michael Scriven (eds.) Perspectives of Curriculum Evaluation. AERA Monograph Series on Curriculum Evaluation, no. 1, Chicago: Rand McNally and Co., 1967, 39-83.

Steinbruner, John O. The Cybernetic Theory of Decision. Princeton: Princeton University Press, 1974.

Suchman, Edward A. Evaluative Research: Principles in Public Service and Action Programs. New York: Russell Sage Foundation, 1967.

Thompson, James D. Organizations in Action. New York: McGraw Hill, 1967.

Tripodi, Tong, Phillip Fellin and Irwin Epstein. Social Program Evaluation: Guidelines for Health, Education, and Welfare Administrators. Itasca, Illinois: F.E. Peacock Publishers, Inc., 1971.

Weidman, Donald R., Pamela Horst, Grace M. Taher, and Joseph S. Wholey. "Design of an Evaluation System for NIMH." Contract Report 962-7. Washington, D.C.: The Urban Institute, (Jan. 15, 1973).

Weiss, Carol H. (ed.) Evaluating Action Programs. Boston: Allyn and Bacon, Inc., 1972.

Weiss, Carol H. Evaluation Research: Methods of Assessing Program Effectiveness. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1972.

Wholey, Joseph S., John W. Scanlon, Hugh G. Duffy, James S. Fikumoto, Leona M. Vogt. Federal Evaluation Policy. Washington, D.C.: The Urban Institute, 1971.

Williams, Walter, and John W. Evans. "The Politics of Evaluation: The Case of Headstart." Annals of the American Academy of Political and Social Science, (September, 1969).

OE 90 8 857